

CAI
BS 113
70 D06

CANADA
DOMINION BUREAU OF STATISTICS
CENSUS DIVISION

Canada

(3)

COMPUTER PROCESSING ERRORS IN THE
1961 CENSUS WAGE-EARNINGS DATA

by

David E. Gower

Working Paper (Demographic and Socio-Economic Series) No. 6



Ottawa, January 1970

31761118484294



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

<https://archive.org/details/31761118484294>

CA1
BS 113
- 70S06

CANADA
DOMINION BUREAU OF STATISTICS
CENSUS DIVISION

Canada

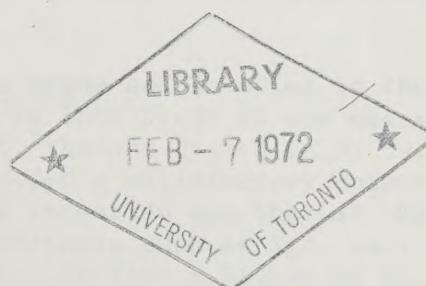
(3)

COMPUTER PROCESSING ERRORS IN THE
1961 CENSUS WAGE-EARNINGS DATA

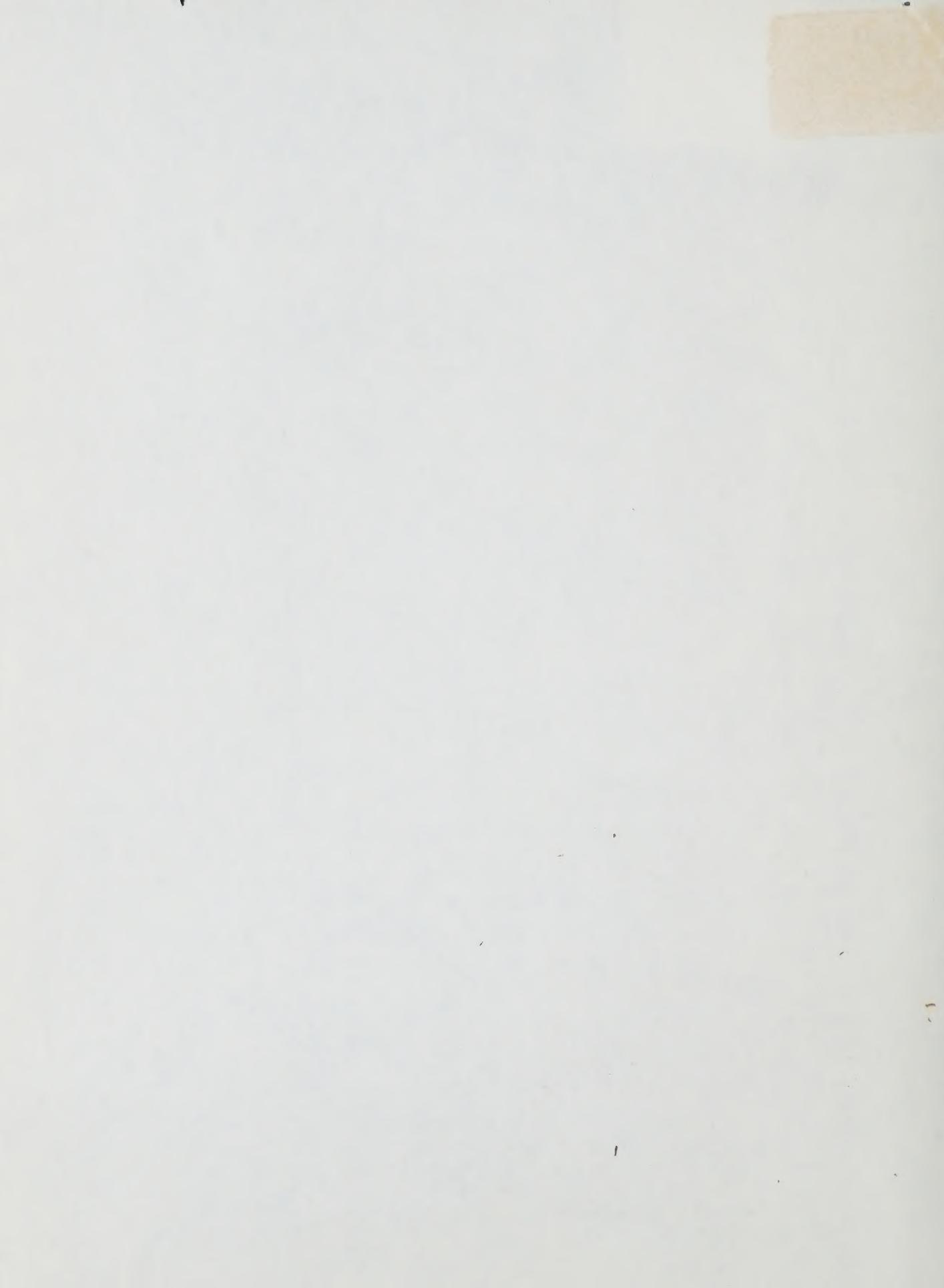
by

David E. Gower

Working Paper (Demographic and Socio-Economic Series) No. 6



Ottawa, January 1970



Computer Processing Errors in the 1961 Census Wage-Earnings Data

1. Introduction

This paper deals with two errors in the computer processing of the 1961 wages and salaries data which have been proven to have resulted in serious distortions of the data. It is hoped that this information will enable interested users to make their own adjustments to 1961 data.

The information contained in this paper has been gathered by the author in the course of two year's work as Head of the Incomes and Earnings Unit of the Economic Characteristics Section, Census Division.

A brief note on the structure of the 1961 Census income data is in order. A question on income from wages and salaries for the 12 months preceding the Census was asked of all wage-earners through face-to-face enumeration. In addition, a 20% sample of non-farm, non-institutional households was left a self-enumeration form in which all persons over the age of 15 were asked questions on all sources of income except net farming income. In fact, therefore, the question on wages and salaries was asked twice of 20% of the wage-earners in the 1961 Census.

Two fairly serious errors were made in the computer processing of wages and salaries data. One error involved the wages and salaries reported on the sample document, the other as gathered on a 100% basis.

The root causes of the errors lie with the fact that 1961 was Canada's first computerized Census, and much had to be learned of the difficult art of communications between subject-matter specialists and computer programmers.

2. Error in assignment of Blanks

The less serious of the two computer processing errors affected the wages and salaries data from the sample. It did not affect the wages and salaries data gathered on the 100% document (as published, for instance, in the 1961 volume 3.3).

The problem arose when the wages and salaries in the sample was blank. The computer was instructed to look first at any wages reported for the person on the other document, the one filled in by face-to-face enumeration. In many cases this yielded a satisfactory answer. However, frequently it did not. The original intention was that at this point a method of random assignments using criteria such as age, sex, education etc. would be used for these persons. In fact a figure of \$3,300 was inserted regardless of sex, age or any other characteristic.

This error affected the records of about 48,000 persons (weighted up from the sample), or about 1% of wage-earners. The national aggregate of wages and salaries was not altered appreciably since \$3,300 was very close to the average reported wage. For males, however, this figure was about \$400 below the reported average, and for females \$1,300 above the reported average.

Evidence exists that non-responders to the wages question tend to have slightly lower wages than responders.(1) Therefore, the figure of \$3,300 may be fairly appropriate for males. For females however, it represents a gross overstatement. In bulletin 4.1-2, tables B1 and B4, for instance, the average wages for females is published as \$2,051. Even ignoring downward bias due to non-response, the use of \$3,300 for an assignment of blanks means that the published averages earnings figure is \$13 too high.

A particular subgroup for which large distortion resulted was the category "occupation not stated". This group included about 34,000 of the 48,000 wage-earners affected by the error. As inspection of line 63, table B4-19 and line 27, table B5-11, bulletin 4.1-2, 1961 Census shows, a huge concentration in the \$3,000-3,499 group resulted. Average incomes for this group were distorted in the same manner as for all wage-earners described above, although the effect was much greater. The "true" average for females, for instance, is at least \$500 below the \$2,649 published.

3. Error in Computation of Averages

A much more serious mistake occurred with respect to the wages data gathered from all wage-earners by face-to-face enumeration. Size of earnings was reported in income intervals on the questionnaire, and for calculation of averages each interval was given a weight. Disastrously, the weight assigned to the \$15,000+ class was \$15,000.

The effect on the average wage-earnings was negligible for females and lowered the male average \$77. The effect of this error was negligible on all the lower-income occupations (including nearly all the females) where the numbers of persons reporting \$15,000+ was insignificant. However, for certain occupations the distortion is staggering. Table I shows recalculated averages for the 41 worst-hit occupations, those whose distortion exceeded \$200.

Bluntly put, this error rendered the average earnings data as published, for instance, in volume 3.3 of the 1961 Census, quite useless for inter-occupational comparisons of the higher-earnings occupations for males.

Anyone wishing to make such comparisons, or to compare 1961 and 1971 wage-earnings data, has two choices. He can use 1961 sample data on employment income as published in 1961 Volume 4.1 and the SX series of publications, or he can adjust for himself the 1961 tables which he desires to use.

(1) Gower D. An Evaluation of 1961 Census Wage Data from Canvasser Enumeration and Self-Enumeration, Working Paper (Demographic and Socio-Economic Series) No. 3, Census Division, DBS. Ottawa, March 1968.

The individual user will, in general, lack information on the weight proper for the \$15,000 class, and in most cases he will also lack information on the frequency of this class. The open-end interval published (and, in most runs, tabulated) is \$10,000+.

A user trying to adjust occupational data would probably find Table I useful. However, this would not be helpful for adjusting average earnings for other distributions.

In an attempt to provide the user with a general formula for adjusting tables where \$10,000+ is the largest class, the relationship between the 10,000+ frequency, the 15,000+ frequency, the 15,000+ weight and the distortion among the different occupation groups was examined. It was found that fairly reliable estimates of the absolute distortion could be obtained using only one statistic: the percentage of wage-earners (working for wages and salaries) reporting more than \$10,000. It was discovered that this figure gave a good estimate of both the percentage in the \$15,000+ class and the weight of the class. A formula was derived by simple observation which seems to give decent results:*

$$(1) Y_t = Y_p + T^2 \text{ if } T \leq 10$$

$$(2) Y_t = Y_p + T^2 - 10T + 100 \text{ if } T > 10$$

Where

Y_t = "true" or adjusted average

Y_p = original average as published

T = % of wage earners working for wages and salaries reporting more than \$10,000.

Two warnings should be observed in using this formula. First, it can be expected to work satisfactorily only on income distributions of a fairly "typical" shape. That is, unimodal, skewed, with mode less than median, median less than mean and a long upper-income tail stretching to several times the distance of the median point from the vertical axis. In addition, a smooth profile with no sudden changes in slope is needed.

An example of a group for whom this formula will not work is airline pilots, whose income distribution has a mode of \$16,000-\$17,000, with a sharp drop thereafter and very few persons above the mode.

*Due to data limitations, proper measurements of the variation from "true" values cannot be taken, except for the occupation figures. A better guide is the "shape of curve" criterion described above. If this is met, the formula can certainly be taken as an indicator of magnitude. The author apologizes that time and resource limitations prevented the calculation of more rigorous estimation parameters.

Second, care should be taken not to apply this formula to 1961 wages data gathered from the population sample, for which this adjustment is not needed because incomes were given in digits. This exception therefore applies to all data in the 1961 volume 4.1 and the "SX" series of publications. The formula is only applicable to the 1961 data on wage-earnings of wage-earners, as published in volume 3.3.

The \$15,000+ averaging error affected not only the average earnings of certain occupation groups, but also of certain age and education groups. In table 2 the formula described above has been applied to illustrate the adjustments required to certain published figures distributed by these variables.

This error affected more than the relative figures. It was also responsible for a loss of \$154 million from the aggregate wages and salaries gathered for wage-earners in 1961. Most of this loss was borne by males.

TABLE 1. Average Wage-earnings of Male Wage-earners working 40 - 52 weeks,
 35+ Hours, Before and After Adjustment for \$15,000+ Averaging Error,
 Selected Occupations,* Canada 1961

Occupation	Unadjusted average	Adjusted average	No. persons affected
1. Owners & Managers, N.E.S. in Primary Metal industries	9,363	10,996	1,562
2. Owners & Managers, N.E.S. in Paper and allied industries	9,427	10,930	2,708
3. Owners and Managers, N.E.S. in Chem. & Chem. products	9,069	10,514	2,957
4. Owners and Managers, N.E.S. in mining, quarrying and oil wells	9,407	10,844	2,559
5. Owners and Managers, N.E.S. in misc. manuf. industries	7,997	9,400	1,760
6. Owners and Managers, N.E.S. in tobacco products industries	9,168	10,545	179
7. Judges and Magistrates	10,817	12,132	588
8. Owners and Managers, N.E.S. in Services to Bus. Management	8,825	10,098	2,340
9. Owners and Managers, N.E.S. in Textile Industries	8,525	9,675	1,249
10. Owners and Managers, N.E.S. in Leather Industries	7,850	8,911	588
11. Physicians and Surgeons	7,472	8,464	5,776
12. Owners and Managers, N.E.S. in Printing Publishing etc.	8,037	8,967	2,578
13. Owners and Managers, N.E.S. in Knitting mills	8,027	8,916	343
14. Owners and Managers, N.E.S. in Machinery Industries	8,548	9,412	1,896
15. Owners and Managers, N.E.S. in Forestry and Logging	7,308	8,097	852
16. Owners and Managers, N.E.S. in Electrical Products Industries	8,623	9,380	2,555
17. Owners and Managers, N.E.S. in Transportation Equipment Industries	8,721	9,476	2,160
18. Owners and Managers, N.E.S. in Food and Beverage Industries	7,140	7,870	6,521
19. Lawyers and Notaries	7,905	8,629	2,938
20. Owners and Managers, N.E.S. in Rubber Industries	8,048	8,695	477
21. Owners and Managers, N.E.S. in Finance and Insurance	7,580	8,226	20,363
22. Owners and Managers, N.E.S. in Clothing Industries	7,718	8,334	1,222
23. Owners and Managers, N.E.S. in Metal Manufacturing Industries	8,287	8,895	3,121

(1) In this list are the 41 occupations for which the distortion exceeded \$200. They are ranked in descending order of absolute size of distortion.

Occupation	Unadjusted average	Adjusted average	No. persons affected
24. Owners and Managers, N.E.S. in Petroleum and Coal Industries	8,763	9,349	799
25. Owners and Managers, N.E.S. in Non-Metallic Mineral Industries	7,769	8,310	1,354
26. Owners and Managers, N.E.S. in Wholesale Trade	7,015	7,500	19,029
27. Owners and Managers, N.E.S. in Construction	7,426	7,881	7,895
28. Owners and Managers, N.E.S. in Wood Industries	7,110	7,522	1,852
29. Security Salesman and Brokers	6,225	6,614	3,694
30. Postmasters	8,115	8,474	1,916
31. Advertising Managers	7,816	8,151	2,025
32. Dentists	8,268	8,602	306
33. Owners and Managers, N.E.S. in Transportation, Commercial and other Utilities..	7,139	7,460	15,468
34. Musicians and Music Teachers	5,437	5,717	1,303
35. Aircraft Pilots, Navigators etc.	8,135	8,404	1,593
36. Sales Managers	7,578	7,840	21,019
37. Owners and Managers, N.E.S. in Furniture Industries	7,200	7,462	767
38. Owners and Managers, N.E.S. in Retail Trade	5,557	5,812	31,280
39. Actors, Entertainers and Showmen	4,470	4,680	483
40. Chemical Engineers	8,083	8,236	2,595
41. Owners and Managers, N.E.S. in Health and Welfare Services	6,460	6,663	2,071

Sources: Unpublished data available at the Economic Characteristics Section, Census Division, DBS(2).

(2) To obtain the frequency of the \$15,000+ class, a special tabulation on wage-earners reporting 49-52 weeks, 35+ hours was used. The weight for the \$15,000+ class was obtained from a table of employment income by occupation, for the labour force, tabulated from the population sample.

In addition, there were 17 occupations where the difference between the adjusted and unadjusted was less than \$200 but over \$100, and an additional 14 between \$50 and \$100.

Only one female occupation has a distortion of over \$50, and this is female physicians and surgeons, which is in the \$100-\$200 distortion class.

TABLE 2. Average Earnings of Wage-Earners from Wages and Salaries,
 1961 Census, Unadjusted and Adjusted for 15,000+ Averaging Error
 (Estimated by Formula)

Schooling	Age	Unadjusted(1)	Adjusted
Elementary	Total	2,964	2,964
Secondary	Total	3,911	3,914
	45-54	4,778	4,794
University	Total	5,699	5,819
	25-34	5,408	5,424
	35-44	7,122	7,356
	45-54	7,372	7,817
	55-64	7,031	7,452

(1) As published in Bulletin 3.3-5, Table 17-1, 1961 Census of Canada.

